

Interpretação de Imagens como Tarefa Cognitiva: uma Abordagem Inicial com Redes Neurais de Aprendizagem Profunda

Cleverson Carneiro¹, Nicolas Mansur Beleski¹, Gustavo Giménez Lugo¹

¹Universidade Tecnológica Federal do Paraná

Resumo. O presente trabalho aborda a funcionalidade de interpretar imagens como uma tarefa cognitiva em agentes. Para verificar alguns dos limites das abordagens consideradas mais efetivas foram efetuados experimentos com redes convolucionais. A principal contribuição, a partir dos resultados, é a constatação e explicitação de possíveis caminhos para elaborar uma arquitetura cognitiva que, embora utilize aprendizagem profunda para aquisição de entradas, o faça sob um arcabouço cognitivo.

1. Introdução

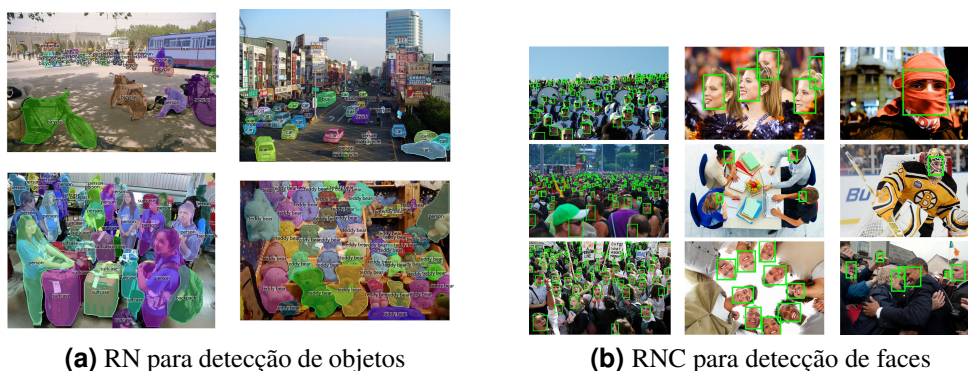
Processamento digital e reconhecimento de objetos em imagens é uma das áreas que mais beneficiou-se com os recentes avanços em inteligência artificial (IA). Este progresso deve-se principalmente aos desenvolvimentos em aprendizagem profunda e redes neurais convolucionais (RNC) [Simonyan and Zisserman, 2014, He et al., 2016].

Apesar destes avanços, ainda existem questões a serem respondidas quanto a capacidade das RNC de generalizarem os conhecimentos obtidos quando comparados, por exemplo, as capacidades humanas de aprendizagem.

Neste trabalho aborda-se a área de pesquisa de ciências cognitivas, que atua buscando compreender como acontece o aprendizado conceitual humano e oferece uma abordagem quanto a construção de modelos que permitem o desempenho humano em tarefas de reconhecimento.

1.1. Limites Teóricos

Apesar de redes neurais atingirem resultados similares a pessoas em alguns desafios de classificação de imagens, a forma como o aprendizado acontece e o conhecimento é processado ainda é bem diferente. A fim de obter nível de resultados de humanos, redes neurais dependem de um treinamento utilizando de dezenas à milhares de imagens, estas que requerem ser manualmente rotuladas para cada classe individual de objeto. Apesar da utilidade destas técnicas em problemas específicos, o extenso treinamento não parece um caminho factível para resolver o problema de classificação em imagens de forma generalizada. Por exemplo, é estimado que pessoas conseguem distinguir aproximadamente 30000 objetos diferentes [Biederman, 1987] e ainda mais categorias que podem ser criadas dinamicamente [Lampert et al., 2014]. Estas diferenças no aprendizado levantam questões quanto possíveis abordagens alternativas para analisar problemas de classificação e reconhecimento em imagens.



(a) RN para detecção de objetos

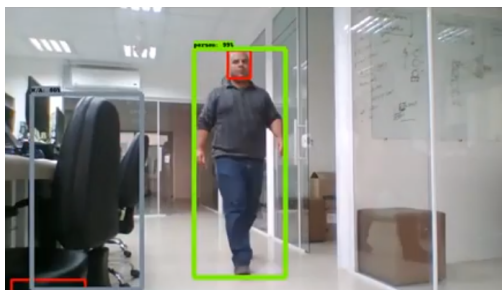
(b) RNC para detecção de faces

Figura 1. Duas implementações utilizadas nos testes.

1.2. Experimento

Para experimentar quanto a eficiência das redes neurais convolucionais no que tange a interpretação de imagem, implementa-se um sistema que busca identificar uma pessoa em uma imagem como passo inicial para uma eventual tarefa do tipo *follow-me*.

O experimento foi realizado usando-se duas redes neurais convolucionais em paralelo (Figura 1 (a) e (b)), combinando-se os resultados individuais utilizando princípios de lógica *naive* baseada em objetivos e utilizando linguagem natural. Ambas redes neurais possuem limitações quanto a detecção de pessoas na forma proposta, entretanto quando aplicadas em conjunto a probabilidade de detecção aumenta, como na Figura 2. A partir destes resultados iniciais, conjectura-se que uma arquitetura que assimile o resultado de múltiplas redes neurais e raciocine sobre estes resultados pode obter soluções mais robustas.

**Figura 2. Combinando redes neurais:** https://youtu.be/_feL5vE24TM

2. Arquitetura Cognitiva

Uma arquitetura cognitiva é composta por sub-sistemas autônomos distribuídos [Fodor, 1983], que percebem seu ambiente, aprendem com a experiência, antecipam os resultados dos eventos, agem para perseguir metas e se adaptam às mudanças das circunstâncias [Ikeuchi, 2014]. Uma arquitetura cognitiva destinada a interpretação de uma imagem pode ser composta pode ser descrita pelas camadas de percepção estruturada, modelo causal e raciocínio.

2.1. Percepção Estrutural

No âmbito do sistema cognitivo visual, [Marr and Vision, 1982] desenvolveu uma teoria computacional no estudo da visão, baseando-se em matemática, psicologia cognitiva, neurociência e estudos clínicos construindo uma hierarquia de diferentes níveis na

estruturação do conhecimento visual. Em [Prinz, 2012] é descrita uma hierarquia organizacional do sistema visual, incluindo as partes do cérebro responsável por tipos diferentes de processamento das informações visuais conforme descrita na Figura 2.1.

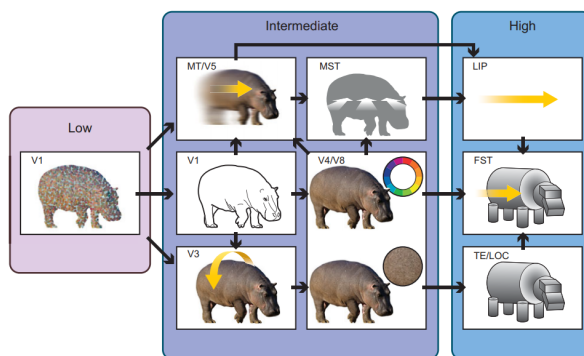


Figura 3. Diagrama do sistema visual [Prinz, 2012]

Se considerarmos as observações feitas por [Marr and Vision, 1982] que separou as regras computacionais de sua implementação e observarmos o trabalho de [Prinz, 2012] que segmentou o processo cognitivo da visão em regiões específicas do cérebro é acreditase que pode-se reproduzir o processo cognitivo da visão através de uma combinação de Redes Neurais Artificiais Especializadas.

3. Representação do Conhecimento

A interpretação de uma imagem é totalmente influenciada pelo propósito atual do agente. Por exemplo, uma cena composta por uma sequência de imagens faz com que haja uma overdose de percepções, que se o agente não tiver algumas restrições prévias definidas em sua base de conhecimento, não trabalhará de forma eficiente.

Para continuar a explorar e testar a arquitetura cognitiva como solução computacional para interpretação de imagens, dá-se ao agente a missão de seguir uma pessoa, desta forma precisar-se representar o conhecimento de "seguir uma pessoa". Segundo [Guarino, 1995] uma maneira de representar o conhecimento é através do uso da ontologia.

3.1. Ontologia

Uma ontologia é uma especificação explícita e formal de uma conceituação [Studer et al., 1998]. Na prática a ontologia é a descrição de classes de um domínio, as propriedades de cada classe descreve seus atributos e regras de restrições dentre as classes e seus atributos. A ontologia, junto com suas instâncias define uma base de conhecimento. Desenvolver uma ontologia inclui [Noy et al., 2001]:

- Definir as classes da ontologia.
- Definir seus atributos e seus valores permitidos.
- Organizar as classes em uma hierarquia taxonômica (sub-classe e super-classe).
- Preencher os atributos das classes para as instâncias.

4. Conclusão

O trabalho realizado constatou que a utilização de redes neurais convolucionais de aprendizagem profunda mostraram-se eficientes em classificar imagens dentro de um cenário restrito e controlado, porém ao aplica-las em imagens parciais, dinâmicas e com baixa nitidez as RNC apresentaram limitações e baixa confiabilidade, principalmente dado a forma de seu treinamento.

A aplicação de múltiplas redes controladas por um componente cognitivo externo, por outro lado, apresentou uma solução mais eficiente. A composição de resultados de diferentes áreas do objeto a ser identificado é inspirada na forma da cognição humana. A aplicação de uma arquitetura cognitiva mostrou-se promissora para aperfeiçoar a performance do agente inteligente.

Inicialmente para a construção da arquitetura cognitiva proposta, verificou-se a importância de formalizar o problema e sua respectiva ontologia. Baseado nos resultados iniciais obtidos, acredita-se que existe espaço e benefícios na abordagem problemas de reconhecimento em imagens utilizando-se de uma perspectiva cognitiva.

Referências

- [Biederman, 1987] Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115.
- [Fodor, 1983] Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- [Guarino, 1995] Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International journal of human-computer studies*, 43(5-6):625–640.
- [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [Ikeuchi, 2014] Ikeuchi, K. (2014). *Computer vision: A reference guide*. Springer Publishing Company, Incorporated.
- [Lampert et al., 2014] Lampert, C. H., Nickisch, H., and Harmeling, S. (2014). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465.
- [Marr and Vision, 1982] Marr, D. and Vision, A. (1982). A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, 1(2).
- [Noy et al., 2001] Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.
- [Prinz, 2012] Prinz, J. (2012). *The conscious brain*. Oxford University Press.
- [Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [Studer et al., 1998] Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2):161–197.